

PERBANDINGAN MODEL ARIMA PADA DATA SPASIAL TRAFIK INTERNET AGREGAT

Sis Soesetijo¹⁾, Febrianto Budimulyono²⁾, Lukas Hadi Purnama³⁾,
Welly Wellandow Santoso⁴⁾, Hendrik Setiawan⁵⁾

^{1,2,3,4,5)} Jurusan Teknik Elektro Universitas Surabaya, Surabaya 60293

email: ¹⁾ ssoesetijo@ubaya.ac.id, ²⁾ febrianto_b4130n@hotmail.com, ³⁾ squall_fire2000@yahoo.com,
⁴⁾ w_ly_inzsight@hotmail.com, ⁵⁾ hendriksetiawan@gmail.com

Abstrak

Pengukuran data spasial trafik internet dilakukan secara agregat selama 46 hari dengan mengambil 4 lokasi pengukuran trafik internet di kampus Universitas Surabaya yaitu Fakultas Bisnis dan Ekonomika, Fakultas Teknik, Perpustakaan dan Kampus Ubaya Ngagel. Pemodelan trafik ini merupakan model trafik internet harian menggunakan model ARIMA (Auto Regressive Integrated Moving Average) dengan validasi model menggunakan qqplot dan uji distribusi normal pada residu model. Oleh karena trafiknya merupakan trafik harian, maka terdapat 184 model ARIMA pada ke-empat lokasi pengukuran trafik tersebut. Hasil pertama yang diperoleh bahwa model ARIMA(1,1,2) merupakan model ARIMA yang umum (sering muncul) pada pemodelan di empat lokasi trafik dan hasil kedua menunjukkan bahwa data spasial trafik internet agregat merupakan data non-stasioner.

Keyword : trafik internet agregat, model ARIMA, non-stasioner, data trafik spasial

1. PENDAHULUAN

Perkembangan trafik internet di Indonesia selama dekade terakhir terbilang sangat tinggi. Bahkan para operator telekomunikasi baik operator telepon tetap maupun bergerak, saat ini lebih cenderung menjual layanan data (internet) daripada layanan suara. Untuk mengetahui karakteristik dan kinerja trafik internet secara lebih komprehensif harus dilakukan pemodelan terhadap data trafik internet sehingga model trafik tersebut dapat digunakan untuk evaluasi, optimasi dan prediksi trafik internet.

Salah satu pemodelan yang paling banyak digunakan untuk pemodelan trafik internet untuk data univariate adalah model ARIMA seperti makalah (Dong, 2005)(H. Zare, 2008)(Huda, 2007). Pada makalah (Dong, 2005) pemodelan ARIMA untuk trafik internet dengan menggunakan data trafik internet secara agregat selama periode pengukuran 14 hari. Sedangkan pada makalah (H. Zare, 2008) periode pengukuran trafik internet agregat hanya 15 menit dengan membedakan antara trafik internet yang terserang hacker (virus) dengan yang tidak. Pada makalah (Huda, 2007), pemodelan trafik internetnya menggunakan model ARIMA yang lebih sederhana dan cepat yaitu model Adjusted ARIMA. Ketiga model di atas merupakan model ARIMA hanya untuk data trafik internet univariate saja.

Sedangkan penelitian pemodelan trafik internet sebagai fungsi multivariate spasial dan temporal telah dilakukan dan dipresentasikan pada makalah (Kensuke, 1999)(Luciano, 2004)(Ioannis, 2009). Makalah (Kensuke, 1999) menggunakan data trafik Round Trip Time (RTT) pada beberapa router dan dianalisis menggunakan spasial korelasi sedangkan pada makalah (Luciano, 2004) menggunakan data trafik TCP dan UDP pada beberapa router. Pada makalah (Ioannis, 2009) membahas pemodelan trafik internet TCP spasial temporal menggunakan model Hidden Markov untuk mendeteksi anomali trafik internet pada jaringan internetnya.

Pada penelitian ini dibahas pemodelan deret waktu univariate trafik internet agregat pada data trafik internet yang diukur secara spasial selama periode 46 hari pada bulan Oktober - Nopember 2010 menggunakan model Auto Regressive Integrated Moving Average (ARIMA). Pengukuran trafik internet dilakukan di kampus Universitas Surabaya (Ubaya) dan secara agregat (pengukuran trafiknya dilakukan tanpa memilih dan memilah protokol trafik internet yang lewat). Untuk mewakili fungsi spasial, pengukuran trafik internetnya dilakukan pada 4 lokasi yaitu trafik internet pada router di fakultas teknik (E), fakultas bisnis dan ekonomika (M), perpustakaan (S) dan kampus Ubaya Ngagel (N). Pada setiap lokasi trafik internet yang diukur selama 46 hari menggunakan perangkat lunak Simple Network Management Protocol (SNMP) akan dipilah menjadi data trafik harian dengan periode trafik harian yang diukur dari jam 00.00-24.00 saja. Dengan demikian terdapat 46 data trafik internet harian pada setiap lokasi trafik sehingga akan menghasilkan model ARIMA sebanyak 46 model. Secara keseluruhan akan diperoleh model ARIMA sebanyak 46 dikalikan 4 lokasi pengukuran sehingga diperoleh total model ARIMA sebanyak 184 model. Masing-masing model ARIMA pada setiap lokasi akan dibandingkan dengan model ARIMA pada lokasi trafik yang lainnya.

Pada penelitian ini diperoleh dua hasil yaitu model ARIMA(1,1,2) merupakan model ARIMA yang banyak muncul di ke-empat lokasi trafik internet dan data trafik internet harian merupakan data non-stasioner. Hasil yang kedua bahwa data internet harian merupakan data non-stasioner (ditunjukkan parameter $d = 1$) ternyata sama dengan hasil yang diperoleh pada makalah (Jin, 2001).

2. MODEL ARIMA(p,d,q)

Autoregressive Integrated Moving Average, ARIMA(p,d,q) didefinisikan dalam bentuk persamaan berikut

$$\phi(B)\nabla^d x_t = \theta(B)a_t \quad (1)$$

di mana: x_t menunjukkan runtun waktu awal

a_t menunjukkan runtun waktu white noise

B merupakan operator backward shift

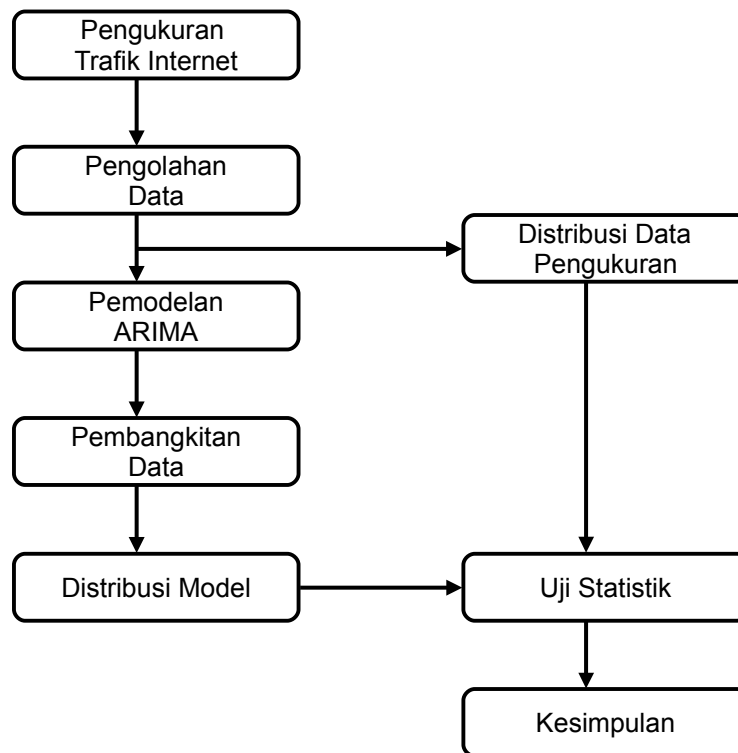
$\nabla^d = (1 - B)^d$ merupakan turunan dengan derajat d

Operator autoregression $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ sedangkan operator moving average $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$. Ketika $d = 0$, persamaan (1) menjadi model ARMA(p,q).

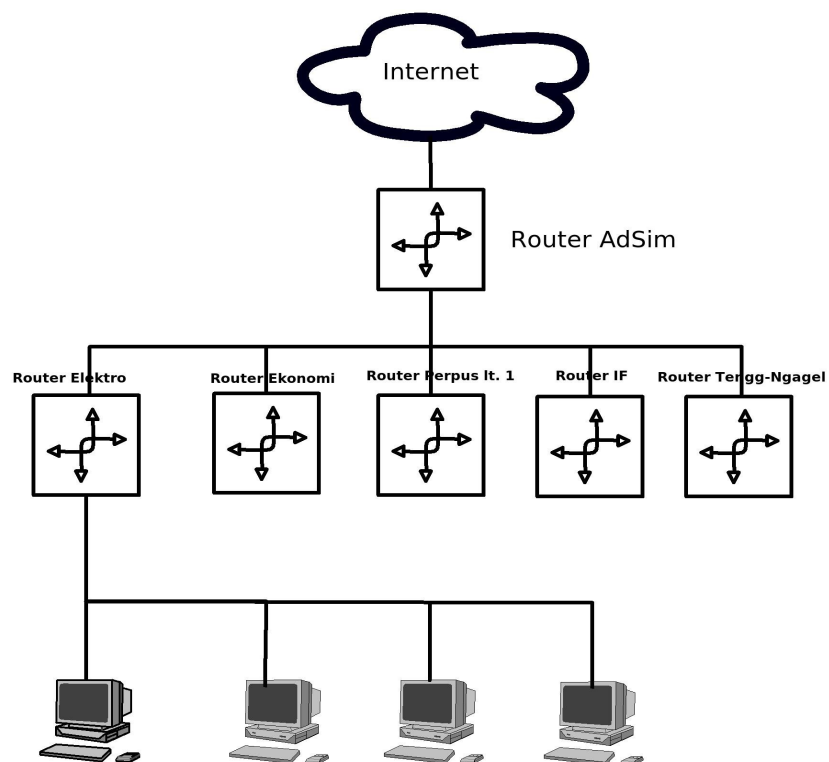
Model ARIMA digunakan secara luas untuk menganalisis data runtun waktu (Dong, 2005). Model ARMA merupakan model runtun waktu univariate untuk data stasioner, sedangkan model ARIMA merupakan model runtun waktu univariate untuk data non stasioner. Proses acak stasioner merupakan karakteristik statistik (seperti rata-rata, variansi, fungsi korelasi) yang dipengaruhi oleh interval waktu. Prinsip dasar model ARIMA adalah mengubah data runtun waktu non stasioner menjadi data stasioner dengan melakukan diferensiasi. Ketika model ARMA(p,q) merupakan model non stasioner, data runtun waktu harus dilakukan penurunan paling tidak sekali untuk mendapatkan model runtun waktu yang stasioner yaitu model ARIMA (p,d,q). Parameter p menyatakan orde dari komponen AR, q menyatakan orde dari komponen MA dan d menyatakan jumlah penurunan yang dilakukan untuk memperoleh data stasioner. Penentuan nilai p dan q tergantung pada fungsi autocorrelation (ACF) dan fungsi partial autocorrelation (PACF). Setelah menentukan orde p dan q dari model ARIMA, kami menggunakan maximum likelihood estimation (MLE) untuk estimasi nilai koefisien ϕ dan θ dan menggunakan Akaike Information Criterion (AIC) untuk mendapatkan parameter yang terbaik (Dong, 2005).

3. METODE PENELITIAN

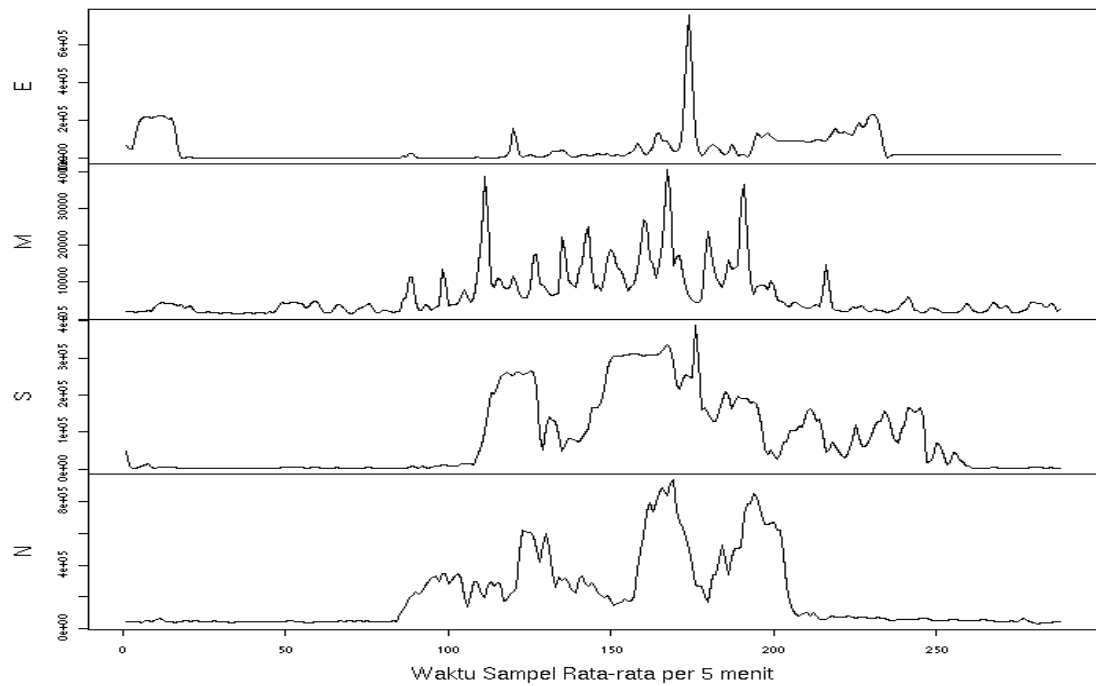
Metode yang digunakan dalam penelitian ini dibagi atas beberapa tahap seperti yang terlihat pada diagram alir metodologi penelitian pada gambar 1 di bawah ini. Langkah awal pada metoda penelitian adalah pengukuran trafik internet secara agregat pada 4 lokasi router di kampus Universitas Surabaya. Keempat lokasi router tersebut adalah router fakultas teknik (E), router fakultas bisnis dan ekonomika (M), router perpustakaan (S) dan router kampus Ubaya Ngagel (N). Pemilihan keempat router diasumsikan dapat mewakili fungsi spasial dari keseluruhan trafik internet di kampus Ubaya. Topologi jaringan internet di kampus Ubaya ditunjukkan pada gambar 2. Untuk mendapatkan data trafik internet agregat harus dilakukan pengambilan data trafik dari router dengan menggunakan Simple Network Management Protocol (SNMP). Pengambilan data trafik internet menggunakan SNMP yang tersimpan dalam bentuk format logfile (Sis, 2011). Data trafik diambil selama 46 hari dari bulan Oktober sampai bulan Nopember 2010. Trafik internet agregat yang diukur merupakan trafik *downstream* saja. Pada setiap lokasi trafik internet yang diukur selama 46 hari dipilah menjadi data trafik harian dengan periode trafik harian yang diukur dari jam 00.00-24.00 saja. Pengukuran trafik internet harian menggunakan rata-rata trafik setiap 5 menit (5 minutes average). Hasil pengukuran trafik internet dalam bentuk runtun waktu untuk data trafik internet harian tanggal 14 Oktober 2010 ditunjukkan pada gambar 3 berikut. Dengan demikian terdapat 46 data trafik internet harian pada setiap lokasi trafik sehingga akan menghasilkan model ARIMA sebanyak 46 model. Secara keseluruhan akan diperoleh model ARIMA sebanyak 46 dikalikan 4 lokasi pengukuran sehingga diperoleh total model ARIMA sebanyak 184 model. Masing-masing model ARIMA pada setiap lokasi akan dibandingkan dengan model ARIMA pada lokasi trafik yang lainnya. Setelah semua proses pengambilan dan pengolahan data selesai, kemudian dilakukan pemodelan ARIMA(p,d,q) menggunakan perangkat lunak statistik open source statistik R. Sebelumnya data diuji terlebih dulu apakah sudah stasioner atau belum. Langkah ini perlu dilakukan untuk menghindari model yang bersifat heteroskedatik. Untuk menguji apakah model yang dihasilkan sudah sesuai maka dilakukan validasi model dengan menggunakan metoda qqplot yaitu perbandingan distribusi model dengan data pengukuran data trafik dan uji distribusi normal pada residu model. Semua perhitungan dan analisis data trafik pada penelitian ini menggunakan perangkat lunak statistik R versi 2.12.2 pada sistem operasi Mandriva linux.



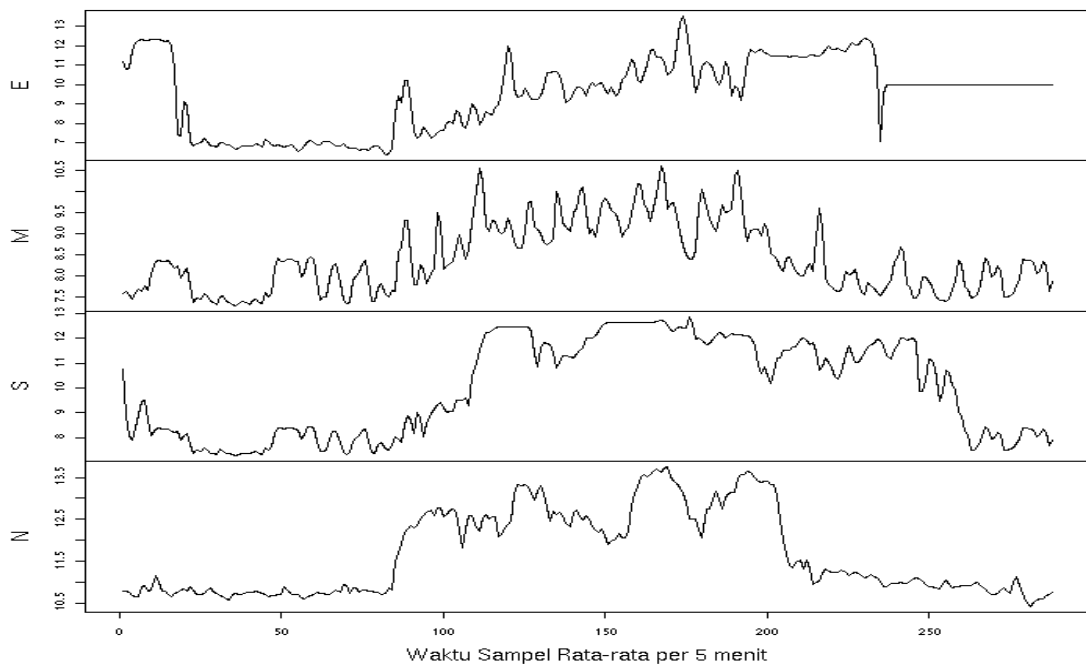
Gambar 1. Metodologi Penelitian



Gambar 2. Topologi Jaringan komputer di Kampus Ubaya



Gambar 3. Grafik Runtun Waktu Trafik Internet Harian Tanggal 14 Oktober 2010



Gambar 4. Grafik Runtun Waktu Trafik Internet Harian Setelah dilognaturalkan

Namun sebelum dimodelkan, perlu proses lognatural pada data trafik internet awal untuk menghindari model yang bersifat heteroskedatik (variansi tidak konstan)(Sis, 2011). Gambar 4 menampilkan grafik runtun waktu trafik harian tanggal 14 Oktober 2010 setelah melalui proses lognatural data trafik awal.

Setelah proses tersebut data diuji sifat stasionernya dengan menggunakan metode Augmented Dickey Duller (ADF), hasilnya ditunjukkan pada tabel 1 di bawah ini. Nampak bahwa data trafik internet harian tanggal 14 Oktober 2010 setelah dilognaturkan menunjukkan sifat non-stasioner. Hal ini dibuktikan dari hasil uji pada nilai mutlak test-value lebih kecil dari nilai mutlak critical-value baik pada confidence interval 90% dan 95%. Namun hasil berubah menjadi data stasioner setelah diturunkan satu kali ($d = 1$), hasil ujinya ditunjukkan pada tabel 1, di mana nilai mutlak test-value (14.07) lebih besar dari nilai mutlak critical-value baik pada confidence interval 90% (1.62) dan 95% (1.95).

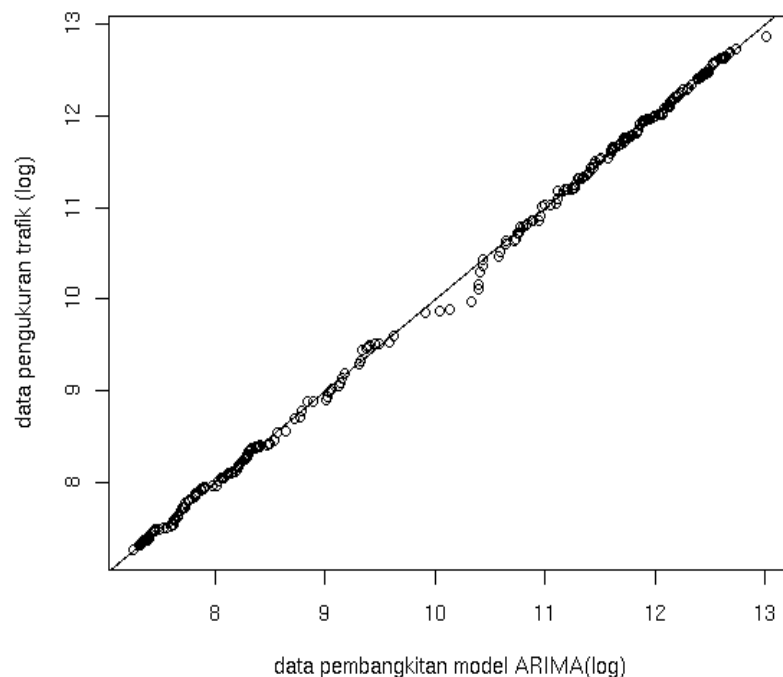
Tabel 1. Hasil Uji Stasioner

Pra-uji	<i>Test-value</i>	<i>Critical Value</i>	
		5 %	10 %
<i>log</i>	-0.39	-1.95	-1.62
<i>diff</i>	-14.07	-1.95	-1.62

Langkah berikutnya adalah menentukan orde p dan orde q dengan menggunakan perhitungan ACF dan PACF. Dari analisis dan perhitungan tersebut diperoleh bahwa orde $p = 2$ dan $q = 1$, sehingga diperoleh model ARIMA (2,1,1). Dengan demikian data trafik internet harian tanggal 14 Oktober 2010 didapatkan model ARIMA (2,1,1) dengan persamaan sebagai berikut :

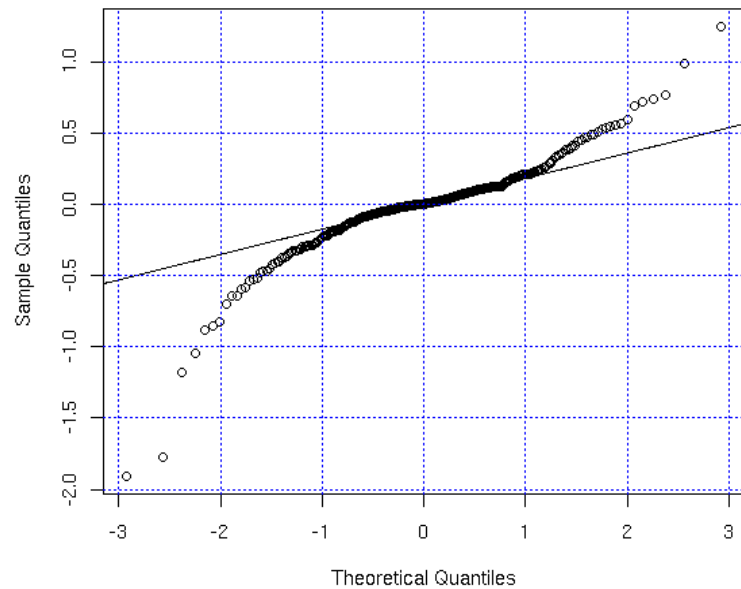
$$Y_t = 0.6252Y_{t-1} - 0.35Y_{t-2} - 0.33\epsilon_{t-1} + e_t \quad (2)$$

Model ARIMA (2,1,1) pada persamaan (2) diperoleh menggunakan metode AIC untuk mendapatkan nilai koefisien terbaik. Dari persamaan (2) di atas kemudian dibangkitkan data trafik untuk dibandingkan distribusinya dengan distribusi data pengukuran trafik harian. Perbandingan distribusi data trafik harian menggunakan metode qqplot. Hasil perbandingan distribusi ditunjukkan pada gambar 5 berikut. Nampak bahwa distribusi data trafik sudah mengikuti garis linier 45 derajat yang berarti distribusi data trafik adalah sama. Hal ini membuktikan bahwa model ARIMA(2,1,1) merupakan model yang valid.



Gambar 5. Perbandingan Distribusi Trafik dengan qqplot

Selain validasi dengan perbandingan distribusi dengan qqplot. Kami menguji distribusi dari residu model (e_t) apakah memiliki distribusi normal dan memiliki rata-rata sama dengan nol. Hasil uji tersebut ditunjukkan pada gambar 6, nampak bahwa distribusi residu model sudah mengikuti garis linier dengan rata-rata residu sama dengan nol. Dengan demikian model ARIMA (2,1,1) dengan validasi qqplot dan uji distribusi normal merupakan model yang valid dan sesuai untuk data trafik agregat harian untuk tanggal 14 Oktober 2010.



Gambar 6. Uji Distribusi Normal Pada Residu Model ARIMA (2,1,1)

4. HASIL DAN PEMBAHASAN

Dengan menggunakan metode yang sama, semua data trafik internet agregat harian selama 46 hari pada 4 lokasi pengukuran trafik dihitung dan dianalisis sehingga diperoleh model ARIMA yang valid dan sesuai. Hasil keseluruhan model ARIMA(p,d,q) ditunjukkan pada tabel 2 dan tabel 3 berikut.

Tabel 2. Model ARIMA untuk Fakultas Teknik dan Ekonomi

ARIMA (p,d,q)	Fakultas Teknik		ARIMA (p,d,q)	Fakultas Ekonomi	
	Jumlah model	%		Jumlah model	%
0,1,1	10	21.74	1,1,2	13	28.26
2,1,0	8	17.39	2,1,2	6	13.04
1,1,2	6	13.04	2,1,0	5	10.86

Tabel 3. Model ARIMA untuk Kampus Ngagel dan Perpustakaan

ARIMA (p,d,q)	Kampus Ngagel		ARIMA (p,d,q)	Perpustakaan	
	Jumlah model	%		Jumlah model	%
0,1,2	5	10.87	0,1,1	6	13.04
1,1,2	5	10.87	1,1,1	6	13.04
2,1,0	5	10.87	1,1,2	6	13.04

Tabel 2 dan 3 hanya menampilkan 3 model ARIMA yang memiliki persentasi yang besar saja dari 46 model ARIMA pada 4 lokasi trafik internet. Dari kedua tabel 2 dan 3 menunjukkan bahwa model ARIMA (1,1,2) merupakan model ARIMA yang sering muncul pada ke-empat lokasi trafik. Hal ini menandakan bahwa model ARIMA(1,1,2) merupakan model ARIMA yang umum untuk ke-empat lokasi terutama pada lokasi trafik fakultas Ekonomi dan Management, kampus Ngagel dan Perpustakaan. Untuk lokasi Fakultas Teknik model ARIMA (1,1,2) berada pada posisi ketiga untuk persentasi model. Secara umum dapat dikatakan bahwa perbandingan model ARIMA pada data trafik internet agregat secara spasial menunjukkan model ARIMA(1,1,2) merupakan model umum.

Hasil kedua dari perbandingan model ARIMA terhadap data trafik harian menunjukkan bahwa data trafik bersifat non-stasioner (secara umum). Data trafik yang bersifat stasioner ditunjukkan pada beberapa data trafik internet agregat harian yaitu data trafik harian pada tanggal 17 dan 21 Nopember 2010 untuk kampus Ngagel, tanggal 10 Oktober, 6 dan 28 Nopember untuk Fakultas Teknik, 6 hari untuk fakultas Ekonomi sedangkan pada lokasi perpustakaan sama sekali tidak terdapat data trafik stasioner.

Dari dua hasil tersebut menunjukkan bahwa karakteristik data trafik agregat harian pada 4 lokasi trafik mempunyai karakteristik yang sama.

5. KESIMPULAN

Hasil perhitungan dan analisis terhadap 46 data trafik internet agregat harian pada 4 lokasi router (pengukuran trafik internet secara spasial) menunjukkan bahwa model ARIMA (1,1,2) merupakan model ARIMA umum dan data trafik internet bersifat non-stasioner. Secara garis besar dapat disimpulkan bahwa data trafik internet agregat pada 4 lokasi mempunyai karakteristik sifat yang sama. Kesamaan karakteristik tersebut diduga berkaitan dengan perilaku pengguna ketika melakukan akses internet.

Hasil yang menunjukkan bahwa data trafik internet agregat bersifat non-stasioner memiliki kesamaan dengan hasil pada makalah (Jin, 2001), pada penelitian tersebut trafik internet yang dianalisis hanya trafik HTTP saja. Apabila dianalisis lebih lanjut ternyata trafik internet agregat yang diukur di kampus Universitas Surabaya menunjukkan bahwa trafik HTTP mempunyai persentasi 97% dari keseluruhan trafik agregat yang ada.

UCAPAN TERIMAKASIH

Makalah ini merupakan bagian dari hasil penelitian yang didukung secara penuh oleh LPPM Universitas Surabaya. Untuk itu penulis mengucapkan banyak terimakasih atas dukungan dana dari LPPM Universitas Surabaya.

6. DAFTAR PUSTAKA

- Bernhard Pfaff, 2008, *Analysis of Integrated and Cointegrated Time Series with R*, Springer, Second Edition
- Dong-Yan Zhang, Ming-Zeng Hu, Hong-Li Zhang, Ting-Biao Kang, 2005, *A Network Traffic Model Based On Measurement*, Proc of 4th International Conferences on Machine Learning and Cybernetics, Guangzhou, China, 18-21 August 2005
- H. Zare Moayed, M.A. Masnadi-Shirazi, 2008, *ARIMA Model for Network Traffic Prediction and Anomaly Detection*, IT Symposium, Kuala Lumpur, Malaysia, 26-28 August 2008
- Huda M. A. El Hag, Sami M. Sharif, 2007, *An Adjusted ARIMA Model for Internet Traffic*, AFRICON 2007
- Ioannis Ch. Paschalidis, Georgios Smaragdakis, 2009, *Spatio-Temporal Network Anomaly Detection by Assessing Deviation of Empirical Measures*, IEEE/ACM Transactions On Networking Vol. 17 No. 3, June 2009
- Jin Cao, William S. Cleveland, Dong Lin, Don X. Sun, 2001, *On the Non-stationary of Internet Traffic*, Proceeding of ACM SIGMETRICS 2001, New York, USA
- Kensuke Fukuda, Hideki Takayasu, Misako Takayasu, 1999, *Spatial and Temporal Behavior of Congestion in Internet Traffic*, Fractal Vol. 7 No.1 page 23-31
- Luciano Rodrigues, Paulo Roberto Guardieiro, 2004, *A Spatial and Temporal Analysis of Internet Aggregate Traffic at the Flow Level*, IEEE Communications Society
- Sis Soesetijo, Febrianto Budimulyono, Lukas Hadi Purnama, Welly Wellandow Santoso, Hendrik Setiawan, 2011, *Analisis dan Pemodelan Trafik Internet Spasial*, SITIA 12th 2011, ITS Surabaya